Math. Res. Lett. **13** (2006), *no.* 4, 587–598

SUB-RIEMANNIAN GEOMETRY AND PERIODIC ORBITS IN CLASSICAL BILLIARDS

Yuliy Baryshnikov and Vadim Zharnitsky

Abstract. Classical (Birkhoff) billiards with full 1-parameter families of periodic orbits are considered. It is shown that construction of a convex billiard with a "rational" caustic (*i.e.* carrying only periodic orbits) can be reformulated as the problem of finding a closed curve tangent to a non-integrable distribution on a manifold. The properties of this distribution are described as well as the consequences for the billiards with rational caustics. A particular implication of this construction is that an ellipse can be infinitesimally perturbed so that any chosen rational elliptic caustic will persist.

1. Introduction and main results

Classical plane billiards [1] were introduced by Birkhoff in the beginning of the century, see *e.g.* his book [3] or [2] as a " special but highly typical systems of this sort", where "the formal side, usually so formidable in dynamics, almost completely disappears, and only the interesting qualitative questions need to be considered" [2]. Indeed, in that very paper Birkhoff illustrated his point by applying Poincaré last theorem to find periodic orbits in smooth convex billiards.

Subsequently, the area-preserving map became a basic model for the Hamiltonian systems with two degrees of freedom. However, in many respects the billiard system is a very special type of area preserving maps, which leads to highly nontrivial problems specific to billiards only.

For example, the so-called Birkhoff conjecture, which states that the only integrable billiards with smooth convex boundary are ellipses, has no analog for the general area preserving monotone twist maps. Here the main difficulty is "reading off" the properties of the billiard map from those of the billiard curve.

Similarly, the well known conjecture that periodic orbits in classical billiards constitute the set of measure zero becomes false for the general monotone-twist area-preserving map. Therefore, it is important to develop the tools pertinent to the billiard like problems.

In this article, we introduce a new approach to study billiards with full 1-parameter families of periodic orbits [2]. The study of such billiards is interesting from the viewpoint of both conjectures mentioned above [3]. Indeed, in integrable billiards there are

Received by the editors July 13, 2004. Revised on May 30, 2005.

The second author was supported by NSF grant DMS-0505216.

[1]The classical billiard system consists of a piece-wise differentiable boundary and a point like ball moving along the straight lines between interceptions of the boundary. At the boundary the ball bounces according to the "angle of incidence is equal to the angle of reflection" rule.

[2]Similar formulation has been developed by A. Török [13].

[3]Also, these billiard tables found recently some applications in nonlinear optics[1]

full 1-parameter families of periodic orbits of various types, as opposed to the generic case when there are just two periodic orbits of each type. Regarding the second conjecture, one should demonstrate that there are no clusters of periodic orbits with positive measure, motivating the study of billiards with "many" periodic orbits.

There exist examples of billiards with continuous family of periodic orbits and with non-elliptic boundary. The most well known examples are the curves of constant width, which possess the full family of two-period orbits, see *e.g.* [6] and references therein. In the three period case, an explicit example was obtained in [5], where the boundary curve is found, such that each point on the boundary is a foot-point of a 3-period orbit. However, there remains an open question as to how large is the set of such billiard boundaries. Also, it remained unclear if similar billiard tables can be found to support caustics carrying periodic orbits of arbitrarily large period.

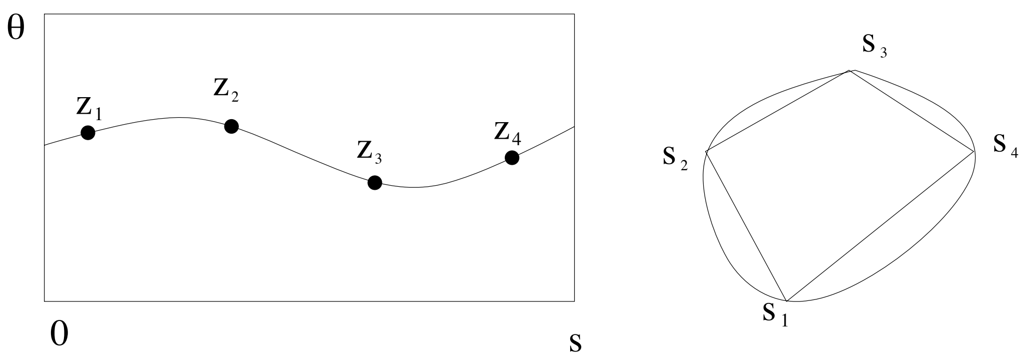

FIGURE 1. Billiard phase space Ω with a continuous family of Birkhoff periodic orbits, s is a natural parameter along the boundary, θ is an angle between the outgoing ray and the tangent measured in the counterclockwise direction.

Here, we address these questions and provide a transparent geometrical description of such billiard domains. First we define formally a rational caustic.

Definition 1.1. *Let* Γ *be a smooth convex billiard boundary. A rational caustic is a curve* γ *possessing the following property: if a billiard orbit segment is tangent to* γ *then so are all other segments of the orbit and the orbit is periodic.*

Recall that the phase space of a billiard can be reduced to $\Omega = \Gamma \times [0, \pi]$, where Γ is the billiard boundary (parameterized by the arclength): the point (s, θ) corresponds to the outgoing ray of the billiard trajectory from point s at the angle θ with the tangent to the boundary. As a dynamical system, the billiard can be described by the self-mapping of Ω and the caustics correspond to the invariant curves of this self-mapping, see Figure 1.

The main tool we use is a new class of (non-holonomic) distributions naturally arising in the billiard problem. To motivate our construction, consider an elliptic billiard, that is the billiard whose boundary Γ is an ellipse. Ellipse is known to possess caustics carrying periodic orbits, *e.g.* 3-period orbits which form triangles (this caustic being an ellipse confocal with Γ). Sliding the triangles around the caustic one obtains a (closed) curve in the space of all triangles. The bisectors of these triangles are orthogonal to the billiard boundary. Then, one can think of the one-parameter family of the triangles as the evolution of a single triangle moving so that its vertices slide orthogonally to the bisectors. In this approach the full family of periodic orbits is a primary object while the boundary Γ is a derived object.

Thus, a 1-parameter family of N periodic orbits defines a curve in a $2N$-dimensional space of plane polygons (corresponding to periodic orbits). The condition that the adjacent sides of the polygon have the same angle with the velocity vector of their common foot translates into the tangency of this curve to an N-dimensional distribution, which turns out to be non-holonomic.

From this viewpoint, the main difficulty is to "close" these orbits. However, this problem turns out to be tractable by the methods of geometric control theory.

Our main result establishes that billiards with rational caustics form a smooth submanifold of finite codimension in a properly chosen Hilbert space. To describe this Hilbert space we need some additional notations.

Let V be the Euclidean plane. Denote by $\sigma : V^N \to V^N$ the linear mapping cyclically permuting points: if $\xi = (\xi_0, \xi_1, \dots, \xi_{N-1}), \xi_k \in V$, then $\sigma\xi = (\xi_1, \xi_2, \dots, \xi_0)$. We will use σ acting on other N-point configuration spaces by cyclically permuting points as well.

Let $I = [0, 1]$. Consider the space $\mathcal{H}$ of H^2-curves[4] $b : I \to \mathbb{T}^N$ in the standard N-dimensional torus satisfying the following *monodromy condition*:

$$\sigma b(0) = b(1). \tag{1}$$

Gluing together N iterations of a curve satisfying the monodromy condition produces a closed curve in $\mathbb{T}^N$. This curve represents a multiple of the diagonal class in 1-homology of $\mathbb{T}^N$ (as the cyclic shift of the components does not change the class). Denote by $\mathcal{H}_\Delta$ the component of $\mathcal{H}$ for which the corresponding homology class is the class of the diagonal.

Intuitively, the points in $\mathbb{T}^N$ correspond to the points of N-periodic orbit on the billiard boundary; the curve in $\mathbb{T}^N$ defines N boundary segments traced by these points. We select the component on which these traces do not overlap, but just match at the endpoints.

Theorem 1.1. *Consider a pair* (Γ, z)*, where* Γ *is a smooth convex billiard boundary possessing a caustic carrying continuous family of convex* N*-period orbits and* z *one of these orbits ("base point"). Assume that the corresponding to* Γ *invariant curve in* Ω *is homotopic to the boundary. Then the space of such pairs close to* (Γ, z) *in* $\mathcal{H} \times V^N$ *forms a Hilbert manifold modeled on a codimension* $(N-1)$*-subspace in* $\mathcal{H} \times \mathbb{R}^N$*.*

A further application of our approach gives another proof that the set of 3-periodic trajectories in a convex smooth billiard is nowhere open, see Section 4.

[4] *i.e.* curves such that the first 2 derivatives are square-integrable. The main reason for using this space is that the billiard boundary will be smooth since the Sobolev embedding theorem implies

$$f \in H^2([0,1]) \Rightarrow f \in C^1([0,1]).$$

2. Birkhoff distribution

2.1. Setup. Now, we give a formal description of our construction. It will be convenient to identify the Euclidean plane V with the complex plane. Consider N distinct ordered points on the plane $\{z_m\}_{m=0}^{m=N-1}$. We will use the convention that the indices form a cyclic group $\mathbb{Z}_N := \mathbb{Z} \text{mod } N$ so that $m+j := (m+j)\text{mod}N$, whenever indices are involved. We consider the set of these N points on the plane as a point in V^N with coordinates $z_m = x_m + \sqrt{-1}y_m, m \in \mathbb{Z}_N$. Let $\mathcal{O}$ be the (open dense) subset of V^N where no three cyclically consecutive point z_{m-1}, z_m, z_{m+1} are collinear. Then N vector fields L_m can be defined over the set $\mathcal{O}$ by

$$L_m z_k = \begin{cases} v_m, & \text{if } k = m, \text{ and} \\ 0 & \text{otherwise,} \end{cases} \qquad m \in \mathbb{Z}_N.$$

where the vectors $v_m \in V$ are defined as

$$v_m := \sqrt{-1}\left[\frac{z_{m+1} - z_m}{|z_{m+1} - z_m|} + \frac{z_{m-1} - z_m}{|z_{m-1} - z_m|}\right].$$

Equivalently, these vector fields in the standard basis are given by

(2)
$$L_m = \left(\frac{y_m - y_{m+1}}{|z_m - z_{m+1}|} + \frac{y_m - y_{m-1}}{|z_m - z_{m-1}|}\right)\frac{\partial}{\partial x_m} + \left(\frac{x_{m+1} - x_m}{|z_m - z_{m+1}|} + \frac{x_{m-1} - x_m}{|z_m - z_{m-1}|}\right)\frac{\partial}{\partial y_m}.$$

Definition 2.1. *The distribution (subbundle of tangent bundle) $\mathcal{B}$ spanned by L_m, $m \in \mathbb{Z}_N$,*

$$\mathcal{B} \subset TV^N \tag{3}$$

is called the Birkhoff distribution*.*

Remark 2.1. *If we provide V^N with the natural symplectic structure $\omega = \sum_i dx_i \wedge dy_i$ (that is sum of the lifts of the symplectic forms $\omega_i = dx_i \wedge dy_i$ on the tangent spaces $T_{z_m}V$), then the Birkhoff distribution is isotropic with respect to ω (i.e. $\omega|_{\mathcal{B}} = 0$). In fact, $\mathcal{B}_z$ is the direct sum of (obviously isotropic) subspaces of $T_{z_m}V$ spanned by v_m, for $m \in \mathbb{Z}_N$.*

If the points $z_m, m \in \mathbb{Z}_N$ taken in their cyclic order form a convex polygon, all frames defined by consecutive pairs v_m, v_{m+1} are all oriented in the same sense (see Figure 2). We will refer to the trajectories having this property as *consistent* trajectories. The trajectories $\{z_o, \ldots, z_{N-1}\}$ forming a convex polygon are consistent. More generally, the trajectories for which $\{z_o, z_k, z_{2k} \ldots, z_{k(N-1)}\}$ form a convex polygon, for some k relatively prime with N are consistent (such trajectories are referred to as *Birkhoff periodic trajectories*).

It is convenient to introduce a basis $\theta = \{\theta_0, \ldots, \theta_{N-1}\}$ for the conormal bundle $N^*\mathcal{B}$ (i.e. subbundle of T^*V^N spanned by the 1-forms vanishing on $\mathcal{B}$) as follows: define θ_k as a linear combination of dx_k, dy_k, $k \in \mathbb{Z}_N$ such that a) $|\theta_k| = 1$, b) $\theta_k(v_k) = 0$ and c) $\theta_k(z_k - z_{k-1}) > 0$ (the first condition is just a normalization; the second one ensures that θ_k is a section of $N^*\mathcal{B}$ and the third condition selects the "outward" direction out of two possible). Remark that condition c) is equivalent also to $\theta_{k-1}(z_{k-1} - z_k) > 0$.

The forms θ can be used to define dually the Birkhoff distribution as the annulator of the bundle spanned by θ.

Define the perimeter function as

$$P = \sum_{m=1}^{N} |z_{m+1} - z_m|. \tag{4}$$

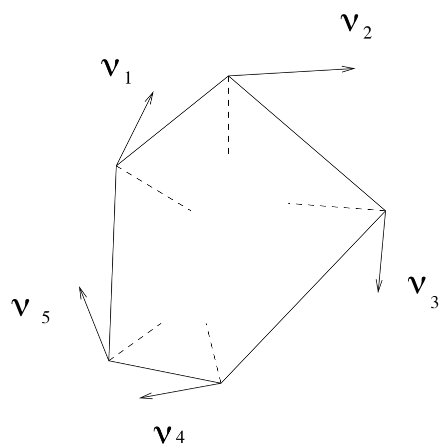

FIGURE 2. The dashed lines are bisectors and the arrows show the vector fields $\nu_1, \nu_2, ..., \nu_5$ which are orthogonal to the bisectors.

It is easy to see that $dP = \sum_k \cos(\alpha_k)\theta_k$, where $2\alpha_k$ is the angle between $z_k - z_{k-1}$ and $z_k - z_{k+1}$.

2.2. Space of billiard trajectories.

Proposition 2.1. *The level set*

$$M_p = \left\{ z \in \mathcal{O} : \frac{z_i - z_{i-1}}{z_i - z_{i+1}} \in \mathbb{C}^1 - \mathbb{R}^1, \ \forall i \in \mathbb{Z}_N; P(z) = p > 0 \right\}$$

is a smooth $2N-1$-dimensional manifold without boundary. Birkhoff distribution is tangent to foliation by M_p's and has rank N over M_p.

Proof: It is easy to check that $dP(z) \neq 0$ for any $z \in M_{P_0}$, which gives the first claim. The second claim follows from the fact that a vertex z_m can only move in the direction orthogonal to the bisector of $\angle(z_{m-1}, z_m, z_{m+1})$ so that the Lie derivative vanishes $L_m P = 0$.

Remark 2.2. *Without loss of generality, we can restrict our attention to M_1 since distributions on M_λ are all equivalent to the one on M_1 by the scaling $(x, y) \to (\lambda x, \lambda y)$. Below, we will omit the subscript in M_1 to avoid cumbersome notation.*

We summarize the informal discussion from the Introduction in the following

Observation: Suppose Γ is a smooth convex boundary possessing full family of periodic orbits. Then these orbits form a curve $z(t) = (z_o(t), ..., z_{N-1}(t)) \in V^N, t \in [0,1]$. This curve $z(t)$ is *horizontal* with respect to the Birkhoff distribution (i.e. $\dot{z}(t) \in \mathcal{B}_{z(t)}, t \in [0,1]$ and satisfies the boundary condition $z_m(1) = z_{m+1}(0)$, for $m \in \mathbb{Z}_N$.

If the invariant curve corresponding to a caustic is homotopic to the boundary in Ω (the only case we are concerned with here), the trajectories forming the curve have nice properties:

Lemma 2.1. *Let Γ be a smooth convex billiard boundary possessing a smooth S^1-parameterized family of $N-$periodic orbits such that the corresponding invariant curve in Ω is homotopic to the boundary. Let $z_i(t), i = 1, ...N$, be the vertices of the periodic orbits parameterized so that $|\dot{z}_1(t)|^2 + ... + |\dot{z}_N(t)|^2 = 1$. Then all the trajectories in the family are Birkhoff periodic and hence consistent and belong to $\mathcal{O}$. Further, all points $z_m, m \in \mathbb{Z}_N$ move in the same direction on Γ, either clockwise or counterclockwise.*

Proof: By a theorem due to Birkhoff, see *e.g. Theorem 12.2.13 in* [7], the invariant curve homotopic to the boundary is a graph in the phase space (see Figure 1). Hence the cyclic order of the points $\{z_o, \ldots, z_{N-1}\}$ does not change within the family, and all the trajectories are necessarily star-like (of the same type (k, N)).

Assume that some of the vertices have zero velocity $\dot{z}_i(t^*)$ at some time $t = t^*$ and $i = i_1, i_2, ...$ Then one of such vertices, say z_j must have a neighboring one with non-zero velocity, say z_{j+1}, for otherwise all vertices would have zero velocity contradicting the assumption. But then the other neighboring vertex z_{j-1} must also possess nonzero velocity which must be pointed counterclockwise for otherwise the Fermat's law at z_j will be violated. But then, at least two vertices move in the different directions which contradicts with the second statement. ☐

2.3. Properties of Birkhoff distribution.

Definition 2.2. *A distribution $\mathcal{D}$ is called bracket generating if the vector fields tangent to $\mathcal{D}$ and all their commutators generate the tangent space T_zM at any point $z \in M$.*

The important property of the Birkhoff distribution is given by

Proposition 2.2. *The Birkhoff distribution is bracket generating of the type $(N, 2N-1)$[5], i.e. the first order commutators already span the $(2N-1)$-dimensional subbundle TM_1.*

Proof of Proposition 2.2: We will prove this statement by explicitly computing commutators of the vector fields $[L_i, L_{i+1}]$[6] and verifying that the set

$$L_1, L_2, ..., L_N, [L_1, L_2], [L_2, L_3], ..., [L_N, L_1] \tag{5}$$

spans the full tangent subspace at any point on M.

In order to organize the calculations, we introduce the notation

$$C_{i,j} = \frac{x_i - x_j}{|z_i - z_j|} \quad S_{i,j} = \frac{y_i - y_j}{|z_i - z_j|} \quad R_{i,j} = |z_i - z_j| \tag{6}$$

[5] See [9].

[6] It is easy to see that $[L_k, L_l] = 0$ if $k - l \mod N \neq \pm 1$.

The following identities are easy to check with direct calculations

$$S_{i,i+1} = -S_{i+1,i} \quad C_{i,i+1} = -C_{i+1,i} \quad S^2_{i,i+1} + C^2_{i,i+1} = 1$$
$$\frac{\partial S_{i+1,i}}{\partial x_i} = -\frac{\partial S_{i,i+1}}{\partial x_i} = \frac{S_{i+1,i}\; C_{i+1,i}}{R_{i+1,i}}$$
$$\frac{\partial C_{i+1,i}}{\partial y_i} = -\frac{\partial C_{i,i+1}}{\partial y_i} = \frac{C_{i+1,i}\; S_{i+1,i}}{R_{i+1,i}}$$
$$\frac{\partial S_{i+1,i}}{\partial y_i} = -\frac{\partial S_{i,i+1}}{\partial y_i} = -\frac{C^2_{i+1,i}}{R_{i+1,i}}$$
$$\frac{\partial C_{i+1,i}}{\partial x_i} = -\frac{\partial C_{i,i+1}}{\partial x_i} = -\frac{S^2_{i+1,i}}{R_{i+1,i}}$$

In (x_i, y_i)-coordinates, the vector fields take the form

$$L_i = (S_{i,i+1} + S_{i,i-1})\frac{\partial}{\partial x_i} - (C_{i,i+1} + C_{i,i-1})\frac{\partial}{\partial y_i}. \tag{7}$$

To simplify the calculations, we choose such coordinates that $y_i = y_{i+1} = 0$ (so that the segment connecting z_i and z_{i+1} lies on the horizontal axis).

Using the above formulae and that $S_{i,i+1} = 0, C_{i,i+1} = -1$, it is then straightforward to compute the commutator:

$$[L_i, L_{i+1}] = \frac{1 - C_{i+1,i+2}}{R_{i,i+1}}\frac{\partial}{\partial x_i} + \frac{C_{i,i-1} + 1}{R_{i,i+1}}\frac{\partial}{\partial x_{i+1}}.$$

Therefore, the commutator $[L_i, L_{i+1}]$ is a vector field *shifting the points* z_i *and* z_{i+1} *along the line* (z_i, z_{i+1}) in such a way that the perimeter does not change.

Now, the computations above imply in fact that

$$\theta_i([L_i, L_{i+1}]) > 0; \quad \theta_{i+1}([L_i, L_{i+1}]) < 0;$$

for all $i \in \mathbb{Z}_N$, and

$$\theta_k([L_i, L_{i+1}]) = 0 \qquad \text{for } k \neq i, i+1.$$

It follows that the matrix $A = (a_{ki})_{k,i=0,\ldots,N-2}$ with entries

$$a_{ki} = (\theta_k([L_i, L_{i+1}]))_{k,i=0,\ldots,N-2}$$

is lower diagonal with positive entries on the diagonal, and hence $(N-1)$ vector fields $[L_i, L_{i+1}], i = 0\ldots, N-2$ already span $(N-1)$-dimensional space $T_zM/\mathcal{B}_z$.

□

Now, we return to billiard curves having a rational caustic, *i.e.* family of periodic orbits parameterized by $[0,1]$ and satisfying the monodromy condition (1). Assume henceforth that the corresponding invariant curve is homotopic to the boundary.

By lemma 2.1 the points $z_i(t)$ move in the same direction along the billiard curve as t varies in $[0,1]$, whence the velocity is a linear combination of the vector fields L_i with all coefficients having the same sign. Such linear combinations satisfy the following important property:

Lemma 2.2. *Let $\alpha, \beta \in \mathbb{R}^N$ and all $\alpha_i > 0$ (or all $\alpha_i < 0$) for $i \in \mathbb{Z}_N$. Then the linear map*

$$\Psi : \mathbb{R}^N \to T_z V^N \text{mod} \mathcal{B}_z,$$

taking $\beta \in \mathbb{R}^N$ to $[L_\alpha, L_\beta](z)\text{mod}\mathcal{B}_z$ has rank $N-1$, where $L_\alpha := \alpha_1 L_1 + \alpha_2 L_2 + ... + \alpha_N L_N; L_\beta := \beta_1 L_1 + \beta_2 L_2 + ... + \beta_N L_N$.

Proof:
We prove the lemma by verifying that

$$[L_\alpha, L_\beta] \in \mathcal{B} \Rightarrow L_\alpha = \lambda L_\beta. \tag{8}$$

Indeed,

$$[\alpha_1 L_1 + ... + \alpha_N L_N, \beta_1 L_1 + ... + \beta_N L_N] =$$
$$(\alpha_1\beta_2 - \alpha_2\beta_1)[L_1, L_2] + (\alpha_2\beta_3 - \alpha_3\beta_2)[L_2, L_3] + ... + (\alpha_N\beta_1 - \alpha_1\beta_n)[L_N, L_1]$$

If $[L_\alpha, L_\beta](z) \in \mathcal{B}_z$, then θ_k vanishes on this vector, and hence

$$(\alpha_k\beta_{k+1} - \alpha_{k+1}\beta_k)\theta_k([L_k, L_{k+1}]) + (\alpha_{k-1}\beta_k - \alpha_k\beta_{k-1})\theta_k([L_{k-1}, L_k]) = 0.$$

As we know, the evaluations $\theta_k([L_k, L_{k+1}]$ and $\theta_k([L_{k-1}, L_k]$ have opposite signs, and hence all brackets

$$(\alpha_k\beta_{k+1} - \alpha_{k+1}\beta_k)$$

have the same sign (or vanish together).

Without loss of generality, assume all brackets are nonnegative: $\alpha_k\beta_{k+1} - \alpha_{k+1}\beta_k \geq 0$ for $k \in \mathbb{Z}_N$. Dividing by $\alpha_k\alpha_{k+1} > 0$, we obtain

$$\frac{\beta_2}{\alpha_2} \geq \frac{\beta_1}{\alpha_1} \geq ... \geq \frac{\beta_2}{\alpha_2},$$

which immediately implies $\beta_i = \lambda\alpha_i$ for all i. In other words, the dimension of $\ker(\Psi)$ is equal to 1. Therefore, the mapping Ψ has rank $N-1$. □

3. Proofs of Theorem 1.1

Let Γ be a smooth convex billiard boundary possessing a caustic carrying continuous family of convex N-period orbits (with corresponding invariant curve in Ω homotopic to the boundary). It is convenient to introduce new coordinates (s, h) in a collar vicinity $U \supset \Gamma$, with s parameterizing Γ and h being the distance to $\Gamma = \{h = 0\}$. We can think of $s(z)$ as the nearest to $z \in V$ point of Γ: s is then well-defined outside of cut-locus of Γ, and thus in some collar neighborhood of Γ.

Let $z^o = z(t), t \in [0, 1]$ be the family of N-periodic orbits tracing Γ. In (s, h) coordinates the trajectory z^o becomes a smooth curve $\zeta^o : [0, 1] \to \mathbb{T}^N \times \mathbb{R}^N$ satisfying the monodromy conditions (1).

We consider $\mathcal{N} = \mathbb{T}^N \times \mathbb{R}^N$ as the space of the trivial bundle $p : \mathcal{N} \to \mathbb{T}^N$. By construction, ζ^o maps to the zero section of p.

Over ζ^o the (pull back of) Birkhoff distribution is transversal to the fibers of p: indeed, the fibers are spanned by the vector fields shifting the points orthogonally to Γ, and the vector fields L_* spanning $\mathcal{B}$ shift the points tangentially to Γ. Hence this transversality persists in some tubular vicinity $\mathcal{V} \times I_c^N$, where $\mathcal{V}$ is a vicinity of the image of ζ^o in $\mathbb{T}^N$ and $I_c = (-c, c), c > 0$ some interval.

Therefore, over $\mathcal{V}$ the Birkhoff distribution defines a *connection* in the vector bundle p, cf [9] (or rather in some tubular vicinity of $\mathcal{V}$). In particular, any curve tangent to

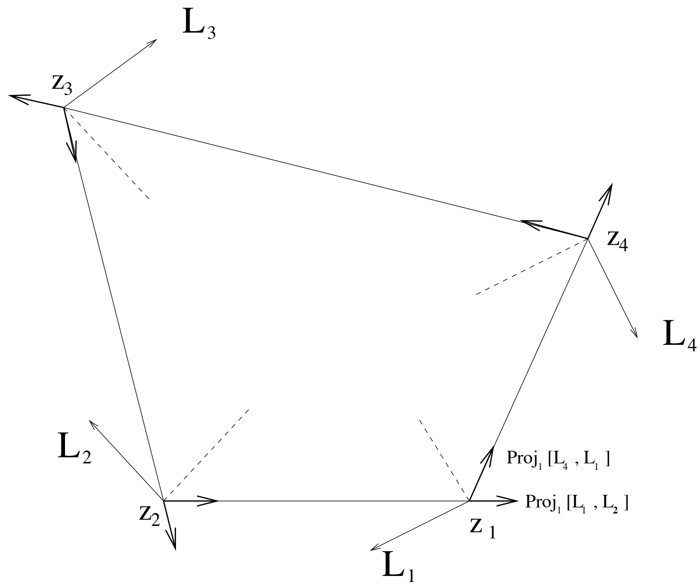

FIGURE 3. The dashed lines are bisectors, and the arrows show the projections of vector fields L_1, L_2, L_3, L_4 which are orthogonal to the bisectors. The short thick arrows show vectors corresponding to commutators.

$\mathcal{B}$ is uniquely defined by a point of the curve and its p-projection: in other words, the curves on the base can be uniquely *lifted* to horizontal (with respect to $\mathcal{B}$) curves, given an initial point. We remark that as ζ^o is horizontal with respect to this connection, all liftings of curves H^2-close to p-projection of ζ^o starting at a point close to $\zeta^o(0)$ will remain close to ζ^o and thus within $\mathcal{V} \times I_c^N$.

As ζ^o satisfies the monodromy condition, so does its projection to $\mathbb{T}^N$, $s_o = p \circ \zeta^o$. The inverse is not true: even if the projection s satisfies $\sigma s(0) = s(1)$, the endpoint $\zeta(1)$ of its horizontal lift does not necessarily match $\sigma\zeta(0)$. On the other hand, if the endpoints match, then $\sigma\dot{s}(0) = \dot{s}(1)$ implies $\sigma\dot{\zeta}(0) = \dot{\zeta}(1)$, as p_* maps planes of $\mathcal{B}$ onto $T_s\mathbb{T}^N$ isomorphically.

For a curve s in $\mathbb{T}^N$ close to s_o in H^2 and satisfying the monodromy conditions, and a point $\zeta(0) \in p^{-1}(s(0))$ close to $\zeta^o(0)$ we define the vector $e(s, \zeta(0)) \in \mathbb{R}^N$ as the difference between $\sigma\zeta(0)$ and the endpoint $\zeta(1)$ of the horizontal lift of s starting at $\zeta(0)$ (both these points belong to the same p-fiber over $s(1)$ which we identify with $\mathbb{R}^N$). Therefore, the space of horizontal curves ζ close to ζ^o and corresponding to the billiard boundaries with rational caustics can be identified with the preimage $e^{-1}(0)$.

Hence, in a standard way, using the implicit function theorem in the theory of calculus in Banach spaces (see *e.g.* [8]), if we could prove that e is a submersion to a manifold of dimension $(N-1)$ (*i.e.* M_1), or, equivalently, that the rank of De is $(N-1)$, the result would follow. This is established below:

Proof of Theorem 1.1 Let $\tilde{L}_k = p_*(L_k), k = 0, \ldots, N-1$ be the projections of the vector fields spanning $\mathcal{B}$ to the base $\mathbb{T}^N$. Fix a smooth bump function $\psi : \mathbb{R} \to \mathbb{R}_+$ with bounded support on the negative half axis. Consider the perturbation of the trajectory $s^o(\cdot) = p(\zeta^o(\cdot))$ in the base near its endpoint $t = 1$. This perturbation is parameterized by $\beta = (\beta_0, \ldots, \beta_{N-1}) \in \mathbb{R}^N$ and defined as

$$s^{\beta,\tau}(t) = s^o(t) + \tau^2 \psi((t-1)/\tau) \sum_k \beta_k \tilde{L}_k(1)$$

for $t \in [0, 1]$, τ small enough (in this formula we use the local affine structure on $\mathbb{T}^N$). Remark that this perturbation leaves s^o unchanged outside of small vicinity of $s^o(1)$ (however, the lift ζ and hence the corresponding family of orbits z are in general perturbed).

The scaling function $\tau^2\psi((t-1)/\tau)$ is chosen to have support in $O(s)$-size vicinity of 1, and to have C^1 norm of order $O(s)$. This ensures that the perturbed curve is H^2-close to s^o.

It is immediate to verify (compare [9]) that the lifts of $s^{\beta,\tau}$ and s^o at 1 differ, up to $O(\tau^4)$ by

$$e_{t_*} := \tau^3[\dot{z}(1), \sum_k \beta_k L_k] + O(\tau^4),$$

where we identified the fibers of p and $T_{\zeta(1)}\mathcal{N}/\mathcal{B}_{\zeta(1)}$. Essentially, the shift of the vertical component under the perturbation is just the value of the curvature form of the connection integrated over the 2-chain bounded by the perturbed and unperturbed curves. The scaling $O(\tau^3)$ is just the total integral of $\tau^2\psi((t)/\tau)$.

Hence, using Lemma 2.2 and the fact that $\dot{z}(t_*)$ expands in L_k with positive coefficients, we conclude that the rank of the mapping $\beta \mapsto e_{t_*}$ is $(N-1)$. Moreover, the parallel transport of the fiber over $s^o(t_*)$ to the fiber over $s^o(1)$ by the horizontal lifting is an isomorphism, and the required result follows.

□

Remark 3.1. *It is worth noticing that to prove this theorem we need just the existence of a single point on the family of N orbits, where the velocities of all vertices are nonvanishing and consistent with an orientation on Γ. Moreover, choosing this point as the basepoint $z^o(0)$, we can derive the existence of a smooth infinite-dimensional family of curves with rational caustics close to Γ and coinciding with Γ outside of arbitrarily small vicinity of the vertices of $z^o(0)$.*

4. Periodic orbits

In this section we apply the above approach from previous sections to prove that 3-period orbits in classical billiards do not contain an open set. This result is originally due to M. Rychlik [11], where a stronger statement is proved that the measure of the set of 3-period orbits is zero. At least three other different proofs appeared subsequently in the literature [12, 15, 14]. We would still like to present another proof to illustrate how the problem can be naturally formulated using the language of nonintegrable distributions. We also hope that this approach may provide new ways of attempting to extend these results to higher period case.

The relation between open set of periodic orbits and the corresponding Birkhoff distribution is given by

Proposition 4.1. *Assume that there exists a smooth convex billiard boundary such that there is an open neighborhood of 3-periodic orbits in the billiard phase space. Then there exists a 2-dimensional integral submanifold* $T\mathcal{D}^2 \in \mathcal{B} \subset TM$.

Proof: To an open set of three period orbits we associate a subset D^2 in M, where M is a manifold of triangles with the unit perimeter. The set $D^2 \in M$ is a submanifold of M. Indeed, Γ is smooth and therefore the map $(s, \theta) \to (z_1, z_2, z_3)$ is smooth, where (s, θ) are the natural parameter along the boundary and the angle of an outgoing ray with the tangent. The map has the full rank (equal to 2). Indeed, let us choose a coordinate system, so that z_1 is at the origin, *i.e.* $x_1 = y_1 = 0$ and that $y_2 = 0$.

Then, we have

$$\frac{\partial x_1}{\partial s} \neq 0, \frac{\partial x_1}{\partial \theta} = 0$$

but it is easy to see that

$$\frac{\partial x_2}{\partial \theta} \neq 0,$$

and therefore the rank of the map is equal to 2.

Now, consider a curve $\gamma(t) \in D^2$, where $t \in [0, \epsilon]$. Then $\dot{\gamma}(0) \in \mathcal{B}$ since $\mathbf{P}_{z_i}\dot{\gamma}(0)$ is tangent to the boundary and therefore orthogonal to the bisector.

□

Now, in order to rule out the existence of an open set of 3-periodic orbits, it suffices to show that there is no 2-dimensional integral manifold in the Birkhoff distribution $\mathcal{B}$ for $N = 3$.

Theorem 4.1. *The Birkhoff distribution with* $n = 3$ *does not admit 2-dimensional integral submanifolds.*

Proof: Suppose that there exists such a submanifold D^2. Then there exist two smooth vector fields X, Y tangent to D^2 and therefore

$$X = a_1 L_1 + a_2 L_2 + a_3 L_3 \tag{9}$$

$$Y = b_1 L_1 + b_2 L_2 + b_3 L_3, \tag{10}$$

where $a_i, b_i \in C^\infty(D^2)$. Without loss of generality, we can assume that $a_1 \neq 0$, then we can modify Y, so that $b_1 = 0$. In this case, either $b_2 \neq 0$ or $b_3 \neq 0$. Again without loss of generality, we can assume that $b_2 \neq 0$ and then we can modify X so that $a_2 = 0$. Furthermore, we can normalize X, Y, so that $a_1 = b_2 = 1$ and then we have

$$X = L_1 + aL_3 \tag{11}$$

$$Y = L_2 + bL_3. \tag{12}$$

Since D^2 is an integral submanifold, then by Frobenius theorem X, Y must be in involution: $[X, Y] \in \{X, Y\}$. We will show that this cannot happen, thus arriving at a contradiction.

Indeed,

$$[X, Y] = [L_1, L_2] - a[L_2, L_3] - b[L_3, L_1]. \tag{13}$$

Using the same argument as in the proof of Lemma (2.2), we find that $[X, Y] \in \mathcal{B}$ only if all coefficients in (13) have the same sign, *i.e.* $a < 0$ and $b < 0$. Let us assume that such a, b are found so that

$$[X, Y] = c_1 L_1 + c_2 L_2 + c_3 L_3 \mod \mathcal{B},$$

where the coefficients also must be positive $c_1, c_2, c_3 > 0$.

Consider now representation of $X, Y, [X, Y]$ in the basis of L_1, L_2, L_3

$$\begin{aligned} \begin{vmatrix} 1 & 0 & c_1 \\ 0 & 1 & c_2 \\ a & b & c_3 \end{vmatrix} &= (c_3 - c_2 b) - c_1 a \\ &= -ac_1 - bc_2 + c_3 \\ &\neq 0, \end{aligned}$$

since $c_1, c_2, c_3 > 0$ and $a, b < 0$. This contradicts our assumption that X, Y are in involution since $X, Y, [X, Y]$ are linearly independent. □

References

[1] Y. Baryshnikov, P. Heider, W. Parz and V. Zharnitsky, *Whispering gallery modes inside the asymmetric resonant cavities*, Phys. Rev. Letters **93** (2004) 4839–4842.

[2] G.D. Birkhoff, *On the periodic motions of dynamical systems'*, Acta Math. **50** (1927) 359–379.

[3] ———, Dynamical systems, American Mathematical Society, Providence, RI, 1966.

[4] B. Halpern, *Strange billiard table*, Trans. Am. Math. Soc. **232** (1977) 297–305.

[5] N. Innami, *Convex curves whose points are vertices of billiard triangles*, Kodai Math. J. **11** (1988) 17–24.

[6] A. Katok and E. Gutkin, *Caustics for inner and outer billiards*, Comm. Math. Phys. **173** (1995) 101–133.

[7] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, (1995) 802 pp.

[8] S. Lang, Differential manifolds, Springer-Verlag, 1985.

[9] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications.* Mathematical Surveys and Monographs, 91. AMS, Providence, RI, 2002.

[10] P. Piccione and D.V. Tausk, *Variational aspects of the sub-Riemannian Geodesic problem*, J. Geom. Phys. **39** (2001) 183–206.

[11] M. Rychlik, *Periodic orbits of the billiard ball map in a convex domain*, J. Differential Geom. **30** (1989) 191–205.

[12] L. Stojanov, *Note on the periodic points of the billiard*, J. Differential Geom. **34** (1991) 835–837.

[13] A. Török, unpublished manuscript.

[14] M. Wojtkovski, *Two applications of Jacobi fields to the billiard ball problem*, J. Differential Geom. **40** (1994) 155–164.

[15] Y. Vorobets, *On the measure of the set of periodic points of a billiard*, Mat. Notes **55** (1994) 455–460.

[16] V. Zharnitsky, *Invariant tori in the Hamiltonian systems with impacts*, Comm. Math. Phys. **211** (2000) 289–302.

Mathematics Research Center, Bell Labs, Lucent Technologies, Murray Hill, NJ 07974
E-mail address: ymb@research.bell-labs.com

Department of Mathematics, University of Illinois, 1409 W. Green Street, Urbana, IL 61801.
E-mail address: vz@math.uiuc.edu